

# A glocal distance for network comparison

G. Jurman, R. Visintainer, S. Riccadonna, M. Filosi, C. Furlanello  
FBK, Trento, Italy

Trento, July 26, 2012

## Introduction

When comparing networks (with the same number of nodes) with direct methods, a number of possible distances is already available in literature. Among others, two of the most common families are the set of edit-like distances and the spectral distances. The functions in the former family quantitatively evaluate the differences between two networks in terms of minimum number of edit operations (with possibly different costs) transforming one network into the other, that is, deletion and insertion of links, while spectral measures relies on functions of the eigenvalues of one of the connectivity matrices of the underlying graph.

A noticeable issue affecting edit distance is the fact of being local, *i.e.* not taking into account the global structure of the networks but only summing the contributions coming from each single link. On the other hand, spectral measures cannot distinguish isomorphic or isospectral graphs. We propose here a possible solution to overcome both issues: combining together an edit (the normalized Hamming) and a spectral distance (the normalized Ipsen-Mikhailov) in a product metric we will denote as HIM. The proposed solution can be applied to any pair of simple networks, *i.e.*, undirected or directed graphs and weighted or unweighted nets, provided the weights lie in the unitary interval  $[0, 1]$ . In what follows we define the two components and the HIM metric itself, with a few examples of applications.

## Notations

Hereafter  $\mathcal{N}_1$  and  $\mathcal{N}_2$  are two simple networks on  $N$  nodes, described by the corresponding adjacency matrices  $A_1$  and  $A_2$ , with  $a_{ij}^{(1)}, a_{ij}^{(2)} \in \mathcal{F}$ , where  $\mathcal{F} = \mathbb{F}_2 = \{0, 1\}$  for unweighted graphs and  $\mathcal{F} = [0, 1]$  for weighted networks. We will moreover denote by  $\mathbb{I}_N$  the identity  $N \times N$  matrix  $\mathbb{I}_N = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$ , by  $\mathbb{1}_N$  the unitary  $N \times N$  matrix with all entries equal to one and by  $\mathbb{0}_N$  the null  $N \times N$  matrix with all entries equal to zero. Moreover, we denote by  $\mathcal{E}_N$  the empty network with  $N$  nodes and no links (with adjacency matrix  $\mathbb{0}_N$ ) and by  $\mathcal{F}_N$  the undirected full network with  $N$  nodes and all possible  $N(N-1)$  links (whose adjacency matrix is  $\mathbb{1}_N - \mathbb{I}_N$ ).

For a directed network  $\mathcal{N}^\uparrow$ , following the convention in [1], a link  $\overset{i}{\bullet} \rightarrow \overset{j}{\bullet}$  is represented by setting  $a_{ji} = 1$  in the corresponding adjacency matrix  $A_{\mathcal{N}^\uparrow}$ , which thus is, in general, not

symmetric. For instance, the matrix  $A_{\mathcal{N}^\uparrow}$  represents the full directed network  $\mathcal{F}_N^\uparrow$ , with all possible  $N^2 - N$  directed links  $\overset{i}{\bullet} \rightarrow \overset{j}{\bullet}$ . The example for  $N = 4$  is shown in Fig. 1.

## Hamming distance

Since computing (plain) edit distance is a NP-hard task, we choose the simplest member of the edit distance family, the Hamming distance, which evaluates the presence/absence of matching links on the two networks being compared.

Hamming distance is one of the most common dissimilarity measures in coding and string theory, and recently it has been used for (biological) network comparison in [2, 3].

The definition of the Hamming distance is the following:

$$\text{Hamming}(\mathcal{N}_1, \mathcal{N}_2) = \sum_{1 \leq i \neq j \leq N} |A_{ij}^{(1)} - A_{ij}^{(2)}|.$$

To guarantee independence from the network dimension (number of nodes), we normalize the above function by the factor  $\bar{\eta} = \text{Hamming}(\mathcal{E}_N, \mathcal{F}_N) = N(N - 1)$ :

$$H(\mathcal{N}_1, \mathcal{N}_2) = \frac{1}{N(N - 1)} \sum_{1 \leq i \neq j \leq N} |A_{ij}^{(1)} - A_{ij}^{(2)}|. \quad (1)$$

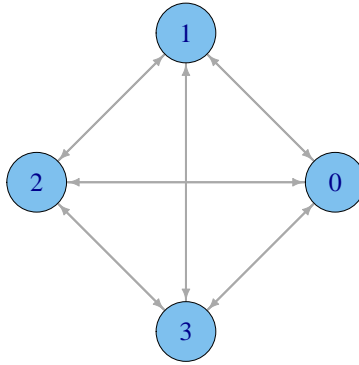


Figure 1: The full directed network  $\mathcal{F}_4^\uparrow$ .

When  $\mathcal{N}_1$  and  $\mathcal{N}_2$  are unweighted networks,  $H(\mathcal{N}_1, \mathcal{N}_2)$  is just the fraction of different matching links (over the total number  $N(N-1)$  of possible links) between the two graphs. In all cases,  $H(\mathcal{N}_1, \mathcal{N}_2) \in [0, 1]$ , where the lower bound 0 is attained only for identical networks  $A_1 = A_2$  and the upper bound 1 is reached whenever the two networks are complementary  $A_1 + A_2 = \mathbb{1}_N - \mathbb{I}_N = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}$ .

## Ipsen-Mikhailov distance

Originally introduced in [4] as a tool for network reconstruction from its Laplacian spectrum, the definition of the Ipsen-Mikhailov  $\epsilon$  metric follows the dynamical interpretation of a  $N$ -nodes network as a  $N$ -atoms molecules connected by identical elastic strings as in Fig. 2, where the pattern of connections is defined by the adjacency matrix of the corresponding network. The dynamical system is described by the set of  $N$  differential equations

$$\ddot{x}_i + \sum_{j=1}^N A_{ij}(x_i - x_j) = 0 \quad \text{for } i = 0, \dots, N-1. \quad (2)$$

We recall that the Laplacian matrix  $L$  of an undirected network is defined as the difference between the degree  $D$  and the adjacency  $A$  matrices  $L = D - A$ , where  $D$  is the diagonal matrix with vertex degrees as entries.  $L$  is positive semidefinite and singular [5, 6, 7, 8, 6], so its eigenvalues are  $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1}$ . The vibrational frequencies  $\omega_i$  for the network model in Eq. 2 are given by the eigenvalues of the Laplacian matrix of the network:  $\lambda_i = \omega_i^2$ , with  $\lambda_0 = \omega_0 = 0$ . In [5], the Laplacian spectrum is called the vibrational spectrum. Estimates (also asymptotic) of the eigenvalues distribution are available for complex networks [9]. Moreover, the relation between the spectral properties and the structure and the dynamics of a network are discussed in [10, 11, 12].

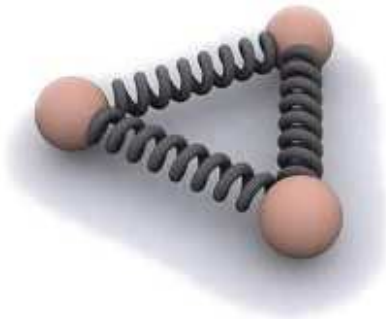


Figure 2: A three nodes network as a oscillatory system.

The spectral density for a graph as the sum of Lorentz distributions is defined as

$$\rho(\omega, \gamma) = K \sum_{i=1}^{N-1} \frac{\gamma}{(\omega - \omega_i)^2 + \gamma^2} , \quad (3)$$

where  $\gamma$  is the common width and  $K$  is the normalization constant defined as

$$K = \frac{1}{\gamma \sum_{i=1}^{N-1} \int_0^\infty \frac{d\omega}{(\omega - \omega_i)^2 + \gamma^2}} , \quad (4)$$

so that  $\int_0^\infty \rho(\omega, \gamma) d\omega = 1$ . The scale parameter  $\gamma$  specifies the half-width at half-maximum, which is equal to half the interquartile range.

In Fig. 7 we show the plot of the Lorentz distribution for two networks.

Then the spectral distance  $\epsilon_\gamma$  between two graphs  $G$  and  $H$  on  $N$  nodes with densities  $\rho_G(\omega, \gamma)$  and  $\rho_H(\omega, \gamma)$  can then be defined as

$$\epsilon_\gamma(G, H) = \sqrt{\int_0^\infty [\rho_G(\omega, \gamma) - \rho_H(\omega, \gamma)]^2 d\omega} . \quad (5)$$

The highest value of  $\epsilon_\gamma$  is reached, for each  $N$ , when evaluating the distance between  $\mathcal{E}_N$  and  $\mathcal{F}_N$ . Defining  $\bar{\gamma}$  as the (unique) solution of

$$\epsilon_\gamma(\mathcal{E}_N, \mathcal{F}_N) = 1 , \quad (6)$$

we can now define the Ipsen-Mikahilov distance as

$$\epsilon(G, H) = \epsilon_{\bar{\gamma}}(G, H) = \sqrt{\int_0^\infty [\rho_G(\omega, \bar{\gamma}) - \rho_H(\omega, \bar{\gamma})]^2 d\omega} , \quad (7)$$

so that  $\epsilon(G, H) \in [0, 1]$  with upper bound attained only for  $(G, H) = (\mathcal{E}_N, \mathcal{F}_N)$ .

A detailed description of the uniqueness of the solution of Eq. 6 is described in Appendix A. Clearly, isospectral networks (and thus also isomorphic networks) cannot be distinguished by this class of measures, so this is a distance between classes of isospectral graphs: although the number of isospectral networks is negligible for large number of nodes [13], their fraction is relevant for smaller networks.

## The HIM distance

Consider now two copies of the space  $\mathbf{N}(N)$  of all simple undirected networks on  $N$  nodes, and endow the first copy with the Hamming metric  $H$  and the second copy with the Ipsen-Mikhailov distance  $\epsilon$ . Then the two obtained pairs  $(\mathbf{N}(N), H)$  and  $(\mathbf{N}(N), \epsilon)$  are metric spaces. Define now on their Cartesian product the  $L_2$  (Euclidean) product metric [14], normalized by the factor  $\frac{\sqrt{2}}{2}$  to set its upper bound to 1. We call this function the HIM

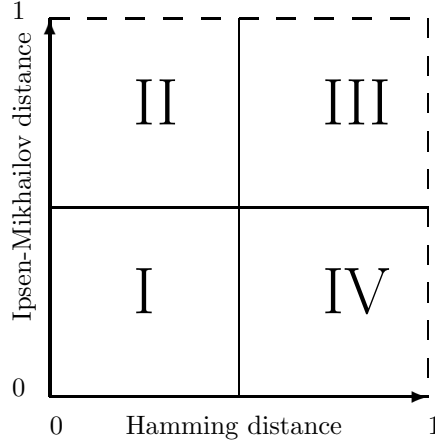


Figure 3: The  $[0, 1] \times [0, 1]$  Hamming/Ipsen-Mikhailov space

metric on the product space, that, with the natural correspondence of the same network in the two spaces, becomes a distance on  $\mathbf{N}(N)$ :

$$\text{HIM}(\mathcal{N}_1, \mathcal{N}_2) = \frac{\sqrt{2}}{2} \|(H(\mathcal{N}_1, \mathcal{N}_2), \epsilon(\mathcal{N}_1, \mathcal{N}_2))\|_2 = \frac{\sqrt{2}}{2} \sqrt{H^2(\mathcal{N}_1, \mathcal{N}_2) + \epsilon^2(\mathcal{N}_1, \mathcal{N}_2)}. \quad (8)$$

The metric  $\text{HIM}(\mathcal{N}_1, \mathcal{N}_2)$  is bounded in the interval  $[0, 1]$ , with lower bound attained for every couple of identical networks, and upper bound attained only on the pair  $(\mathcal{E}_N, \mathcal{F}_N)$ .

Because of the different nature of the two components of the product metric, HIM will be nonzero for non-identical isomorphic/isospectral graphs.

Consider now the  $[0, 1] \times [0, 1]$  Hamming/Ipsen-Mikhailov space: a point  $P(x, y)$  represents the distance of two networks  $G_1, G_2$ , whose coordinates are  $x = H(G_1, G_2)$  and  $y = \epsilon(G_1, G_2)$ , and the norm of  $\vec{P}$  is  $\sqrt{2}$  times the distance  $\text{HIM}(G_1, G_2)$ . If we (roughly) split the Hamming/Ipsen-Mikhailov space into four main zones I, II, III, IV as in Fig. 3, we can say that two networks whose distances correspond to a point in zone I are quite close both in terms of matching links and of structure, while those falling in the zone III are very different with respect to both characteristics. Networks corresponding to a point in zone II have many common links, but their structure is rather different, while a point in zone IV indicates two networks with few common links, but with similar structure.

Examples of couples of networks living in different zones of the Hamming/Ipsen-Mikhailov space are shown in the sections below.

## HIM for directed networks

In case of a pair of directed networks, the corresponding adjacency matrices are not symmetric, and the same happens for the Laplacian matrices. Therefore, their laplacian spectra lie in  $\mathbb{C}$  rather than  $\mathbb{R}$ , and thus computing the Ipsen-Mikhailov distance would require extending

the Lorentzian distribution to the complex plane. A simpler solution can be obtained by transforming the directed network  $D^\uparrow$  into an undirected (bipartite) one  $\hat{D}^\uparrow$ , following the procedure cited in [1]: for each node  $x_i$  in  $D^\uparrow$ , the graph  $\hat{D}^\uparrow$  has two nodes  $x_i^I$  and  $x_i^O$  (where I and O stands for In and Out respectively) and for each directed link  $x_i \rightarrow x_j$  in  $D^\uparrow$  there is a link  $x_i^O - x_j^I$  in  $\hat{D}^\uparrow$ . If the adjacency matrix for  $D^\uparrow$  is  $A_{D^\uparrow}$ , the corresponding matrix for  $\hat{D}^\uparrow$  is  $A_{\hat{D}^\uparrow} = \begin{pmatrix} 0 & A_{D^\uparrow}^T \\ A_{D^\uparrow} & 0 \end{pmatrix}$ , w.r.t. to the node ordering  $x_1^O, x_2^O, \dots, x_n^O, x_1^I, \dots, x_n^I$ .

An example of the above transformation is shown in Fig. 4.

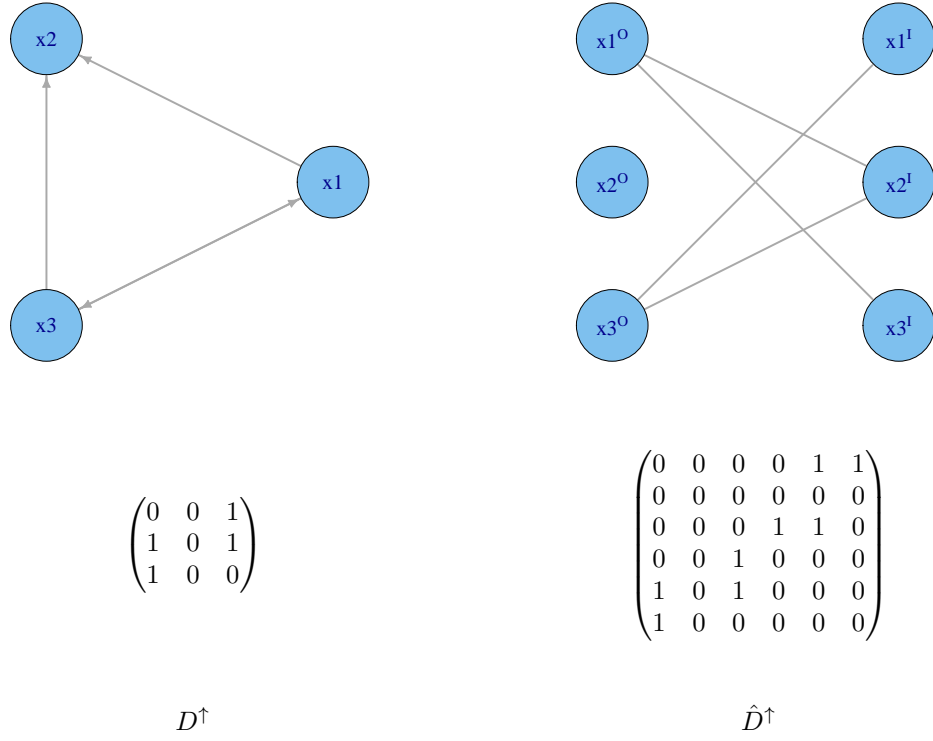


Figure 4: A directed network  $D^\uparrow$  on three nodes and the equivalent undirected network  $\hat{D}^\uparrow$  on six nodes, together with their adjacency matrices.

Thus it is possible to define  $\text{HIM}(\mathcal{N}_1^\uparrow, \mathcal{N}_2^\uparrow)$  as  $\text{HIM}(\hat{\mathcal{N}}_1^\uparrow, \hat{\mathcal{N}}_2^\uparrow)$  after substituting the normal-

izing factors  $\bar{\eta}$  and  $\bar{\gamma}$  with the corresponding  $\bar{\eta}^\dagger$  and  $\bar{\gamma}^\dagger$  derived by imposing the conditions  $\text{Hamming}(\hat{\mathcal{E}}_N, \hat{\mathcal{F}}_N)/\bar{\eta}^\dagger = 1$  and  $\epsilon_{\bar{\gamma}^\dagger}(\hat{\mathcal{E}}_N, \hat{\mathcal{F}}_N) = 1$ , so that  $\text{HIM}(\hat{\mathcal{E}}_N, \hat{\mathcal{F}}_N) = 1$  by using Eq.8. It is immediate to compute  $\bar{\eta}^\dagger = 2N(N-1)$ , while  $\bar{\gamma}^\dagger$  can be numerically computed as for  $\bar{\gamma}$ : details are given in Appendix B.

## Examples of applications

### A minimal example

Consider the two networks  $I_1, I_2 \in \mathcal{N}(8)$  with corresponding adjacency matrices  $A^{I_1}, A^{I_2}$  shown in Fig. 5,6.

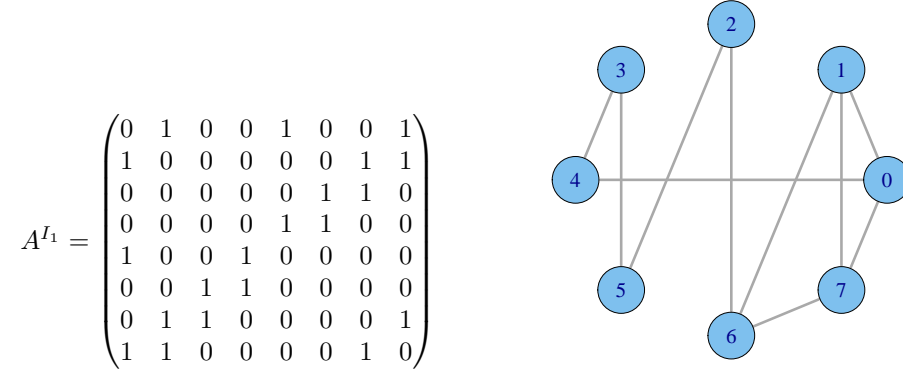


Figure 5: Adjacency matrix and graphical representation of  $I_1$

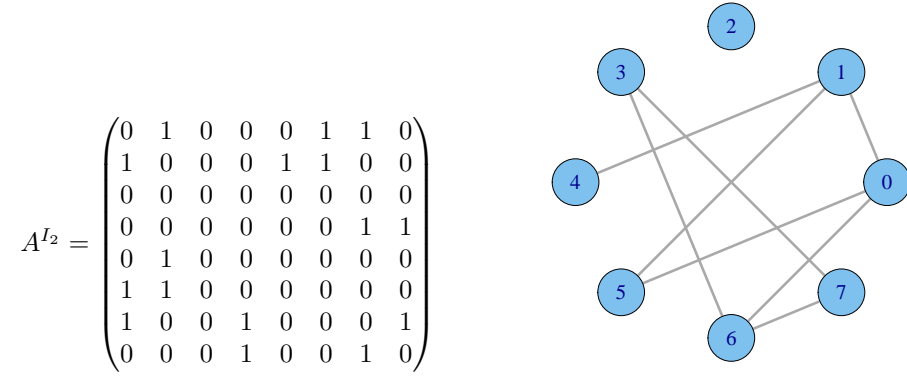


Figure 6: Adjacency matrix and graphical representation of  $I_2$

The Hamming distance between  $I_1$  and  $I_2$  is

$$\begin{aligned} H(I_1, I_2) &= \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} |A_{ij}^{I_1} - A_{ij}^{I_2}| \\ &= \frac{1}{56} \sum_{1 \leq i \neq j \leq 8} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \frac{28}{56} \\ &= 0.5 . \end{aligned}$$

From the spectral point of view, the corresponding Laplacian matrices and eigenvalues



are

$$L^{I_1} = \begin{pmatrix} 3 & -1 & 0 & 0 & -1 & 0 & 0 & -1 \\ -1 & 3 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 2 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 \\ -1 & 0 & 0 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 2 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 3 & -1 \\ -1 & -1 & 0 & 0 & 0 & 0 & -1 & 3 \end{pmatrix} \quad \text{spec}(L^{I_1}) = \begin{bmatrix} 0 \\ 0.657077 \\ 1 \\ 2.529317 \\ 3 \\ 4 \\ 4 \\ 4.813607 \end{bmatrix}$$

$$L^{I_2} = \begin{pmatrix} 3 & -1 & 0 & 0 & 0 & -1 & -1 & 0 \\ -1 & 3 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & -1 & -1 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & 2 \end{pmatrix} \quad \text{spec}(L^{I_2}) = \begin{bmatrix} 0 \\ 0 \\ 0.340321 \\ 1.145088 \\ 3 \\ 3 \\ 3.854912 \\ 4.659679 \end{bmatrix}$$

From the above spectra, we can compute the corresponding Lorentz distributions  $\rho_{I_{\{1,2\}}}(\omega, \bar{\gamma})$ , where  $\bar{\gamma} = 0.4450034$ : their plots are shown in Fig. 7.

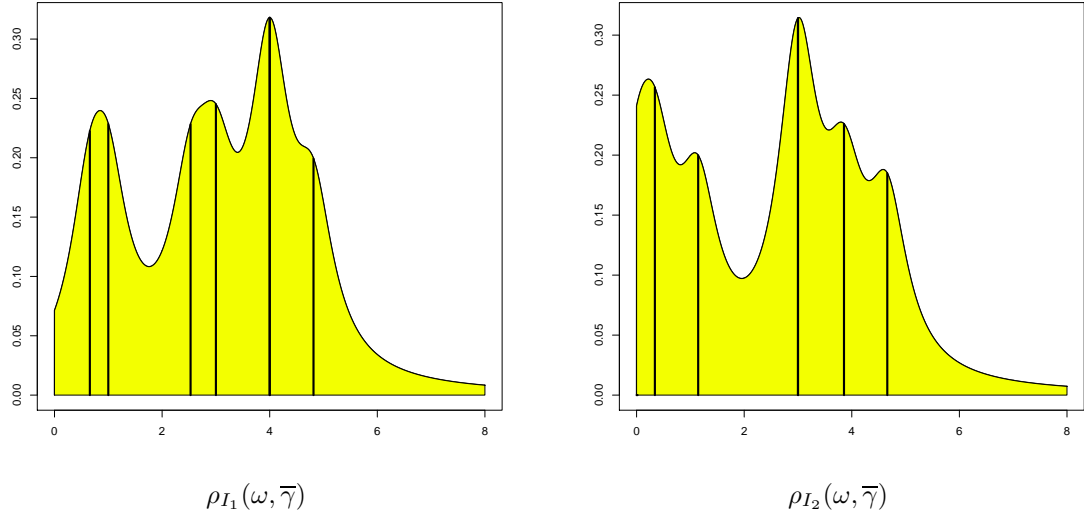


Figure 7: Lorentzian distribution of the Laplacian spectra for  $I_1$  and  $I_2$ . Vertical lines indicate eigenvalues.

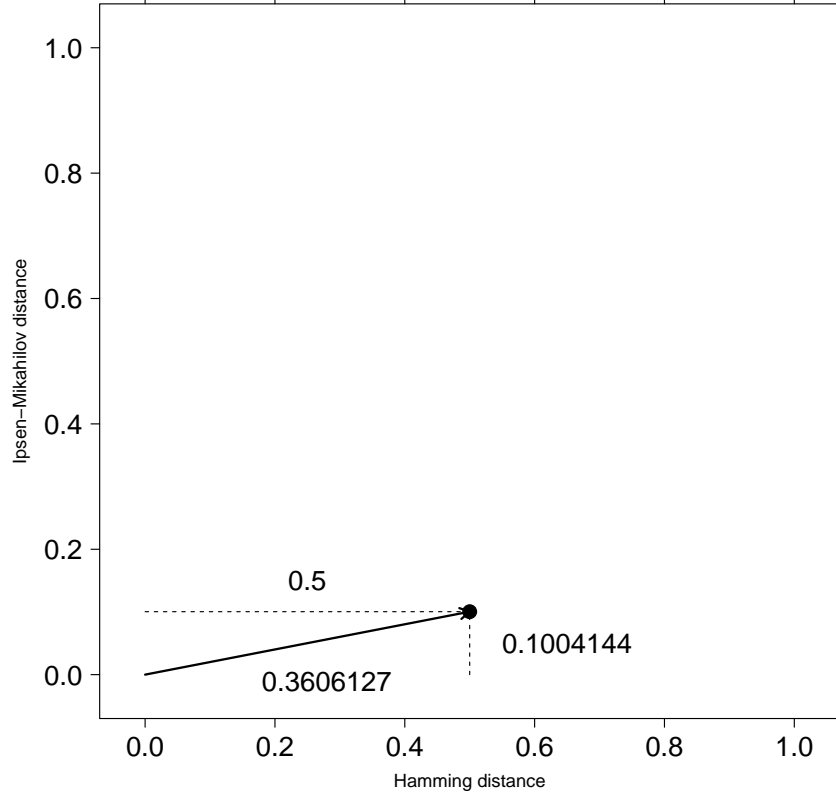


Figure 8:  $\text{HIM}(I_1, I_2)$  in the Hamming/Ipsen-Mikhailov space

The resulting Ipsen-Mikhailov distance is

$$\epsilon(I_1, I_2) = \sqrt{\int_0^\infty [\rho_{I_1}(\omega, \overline{\gamma}) - \rho_{I_2}(\omega, \overline{\gamma})]^2 d\omega} = 0.1004144 ,$$

so that the HIM distance results

$$\text{HIM}(I_1, I_2) = \frac{\sqrt{2}}{2} \|(H(I_1, I_2), \epsilon(I_1, I_2))\|_2 \approx 0.707168 \sqrt{0.5^2 + 0.1004144^2} \approx 0.3606127 .$$

The situation can be graphically represented as in Fig. 8: the two networks are quite different in terms of matching links, but their structures are not so far away.

## A larger study

Fixed the number of nodes  $N$ , there are exactly  $2^{\frac{N(N-1)}{2}}$  different simple undirected unweighted networks, which can be grouped into isomorphism classes. As anticipated before,

isomorphic graphs cannot be distinguished by spectral metrics, while their mutual Hamming distances are non zero, since their links are in different positions. As an example, for  $N = 3$  there are 8 networks grouped in 4 isomorphism classes, for  $N = 4$  there are 11 isomorphism classes including a total of 64 graphs and for  $N = 5$  34 classes with 1024 networks (for  $N = 6, 7$ , the number of classes is respectively 156 e 1044).

To give an overview of a broader situation, we compute a number of mutual distances between networks with a given number of nodes (all possible couples for  $N = 3, 4, 5$  and a subset of them for  $N = 15$ ) and we display the results in Fig. 9. To select a good range of variability for the networks with 15 nodes, we select the empty graph, the full graph (with 105 nodes) and 10 different graphs with  $i$  edges each, for  $1 \leq i \leq 104$ .

As shown by the plots, all possible situations can occur, apart from points in the northwest corner of zone II which are the rarest. For instance, the point  $P(1, 0)$  in Fig. 9(b) corresponds to 6 different pairs  $(O_1, O_2)$  of networks with 4 nodes with maximal Hamming distance and minimal spectral distance: as an example, we show one of these pairs in Fig. 10.

## Biological networks

In [15], the authors used the Keller algorithm to infer the gene regulatory networks of *Drosophila melanogaster* from a time series of gene expression data measured during its full life cycle. They selected 66 time points during the developmental cycle, spanning across four different stages (Embryonic – time points 1-30, Larval – t.p. 31-40, Pupal – t.p. 41-58, Adult – t.p. 59-66), following the dynamics of 588 gene ontological groups and then constructing a time series of inferred networks  $N_i$ <sup>1</sup>. Hereafter we evaluate the structural differences between  $N_i$  and the initial network  $N_1$ , as measured by the HIM distance: the resulting plot is displayed in Fig. 11. The largest variations, both between consecutive terms and with respect to the initial network  $N_1$ , occur in the embrional stage (E). In particular, it is interesting to note that the dynamics of the networks move  $N_i$  away from  $N_1$  until time points 23, then the following terms start getting closer again to  $N_1$  in terms of HIM distance: such behaviour was detected also in the original paper, but only qualitatively, while the introduced metrics can provide a quantitative assessment of the occurring differences.

Finally, it can be appreciated the different range of the two distances: while Hamming distance ranges between 0 and 0.0223, the Ipsen-Mikhailov distance has 0.0851 as its maximum, indicating an higher variability of the networks in terms of structure rather than matching links.

## References

- [1] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.
- [2] K. Tun, P. Dhar, M. Palumbo, and A. Giuliani. Metabolic pathways variability and sequence/networks comparisons. *BMC Bioinformatics*, 7(1):24, 2006.
- [3] E.R. Dougherty. Validation of gene regulatory networks: scientific and inferential. *Briefings in Bioinformatics*, 12(3):245–252, 2010.

---

<sup>1</sup>Adjacency matrices are available at <http://cogito-b.ml.cmu.edu/keller/downloads.html>

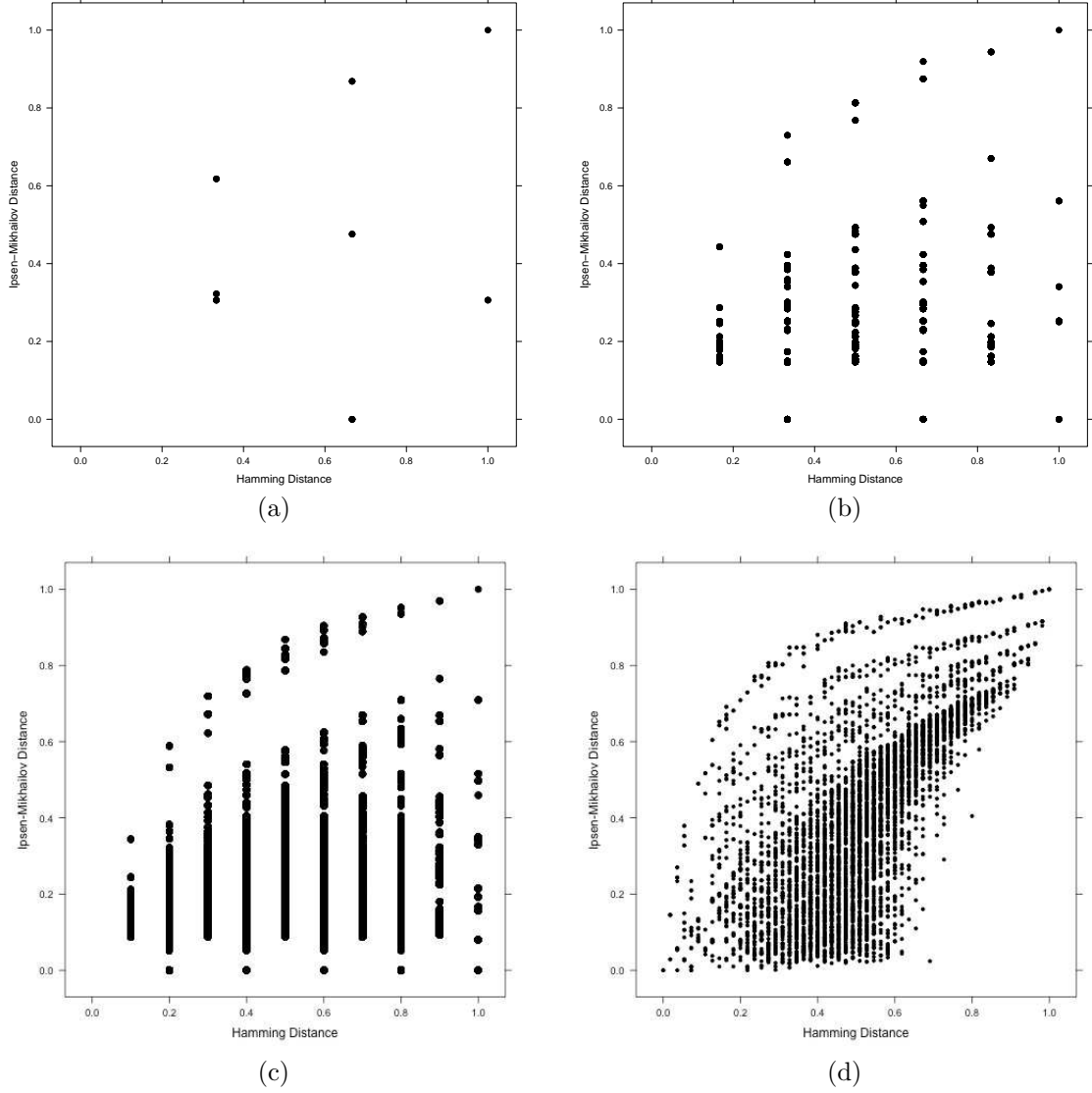


Figure 9: Mutual distances between (a) all 28 couples of networks with 3 nodes, (b) all 2016 couples of networks with 4 nodes, (c) all 523776 couples of networks with 5 nodes and (d) the 542361 mutual distances between a set of 1042 networks with 15 nodes.

- [4] M. Ipsen and A.S. Mikhailov. Evolutionary reconstruction of networks. *Phys. Rev. E*, 66(4):046109, 2002.

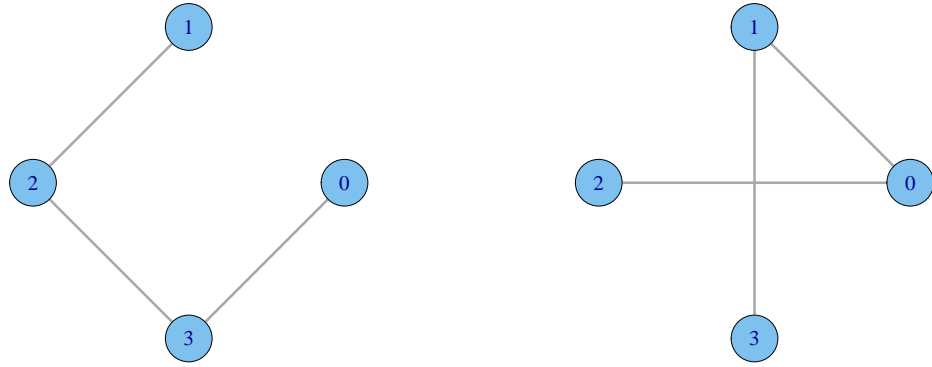


Figure 10: Pair of networks with Hamming distance one and Ipsen-Mikhailov distance zero.

- [5] F. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [6] F.M. Atay, T. Bıyıkoglu, and J. Jost. Network synchronization: Spectral versus statistical properties. *Physica D Nonlinear Phenomena*, 224:35–41, 2006.
- [7] D.A. Spielman. Spectral Graph Theory: The Laplacian (Lecture 2). Lecture notes, 2009.
- [8] R. Tönjes and B. Blasius. Perturbation Analysis of Complete Synchronization in Networks of Phase Oscillators. arXiv:0908.3365, 2009.
- [9] G.J. Rodgers, K. Austin, B. Kahng, and D. Kim. Eigenvalue spectra of complex networks. *Journal of Physics A: Mathematical and General*, 38(43):9431, 2005.
- [10] J. Jost and M.P. Joy. Evolving Networks with distance preferences. *Phys. Rev. E*, 66:036126, 2002.
- [11] J. Jost. Dynamical Networks. In J. Feng, J. Jost, and M. Qian, editors, *Networks: From Biology to Theory*, pages 35–64. Springer-Verlag, 2007.
- [12] J.A. Almendral and A. Díaz-Guilera. Dynamical and spectral properties of complex networks. *New J. Phys.*, 9:187, 2007.

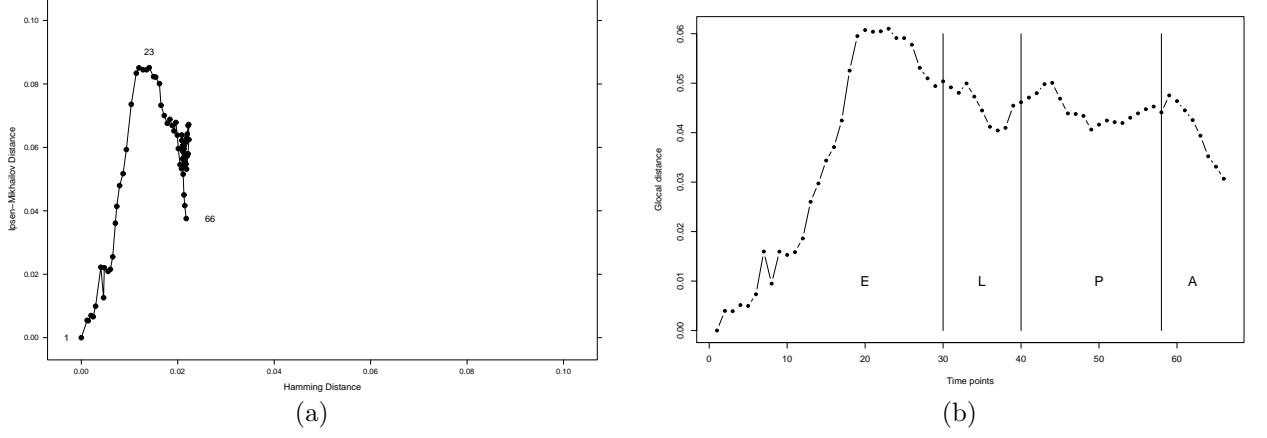


Figure 11: (a) Evolution of distances of the *D. melanogaster* network time series in the Hamming/Ipsen-Mikhailov space and (b) evolution of HIM distances of the *D. melanogaster* network along 66 time points in the 4 stages Embryonic (E), Larval (L), Pupal (P) and Adult (A)

- [13] W.H. Haemers and E. Spence. Enumeration of cospectral graphs. *Eur. J. Comb.*, 25(2):199–211, 2004.
- [14] E. Deza. *Encyclopedia of Distances*. Springer Verlag, 2009.
- [15] M. Kolar, L. Song, A. Ahmed, and E.P. Xing. Estimating time-varying networks. *Ann. Appl. Stat.*, 4(1):94–123, 2010.

## A Uniqueness of $\bar{\gamma}$

Fix the number  $N$  of nodes, and consider the two extremal networks  $\mathcal{E}_N$  and  $\mathcal{F}_N$ , whose Laplacian spectrum is respectively

$$\text{spec}(L(\mathcal{E}_N)) = (\underbrace{0, \dots, 0}_N) \quad \text{and} \quad \text{spec}(L(\mathcal{F}_N)) = (0, \underbrace{N, \dots, N}_{N-1}),$$

so that  $\omega_i = 0$  for the empty network and  $\omega_i = \sqrt{N}$  for the fully connected network, for  $i = 1, \dots, N-1$ .

The Lorentz distribution for the empty network is thus

$$\begin{aligned} \rho_{\mathcal{E}_N}(\omega, \gamma) &= K \sum_{i=1}^{N-1} \frac{\gamma}{\gamma^2 + (\omega - \omega_i)^2} \\ &= \frac{K\gamma(N-1)}{\gamma^2 + \omega^2}, \end{aligned}$$

where  $K$  can be computed as

$$\begin{aligned}
K &= \frac{1}{\int_0^{+\infty} \frac{\gamma(N-1)}{\gamma^2 + \omega^2} d\omega} \\
&= \frac{1}{(N-1) \left[ \arctan\left(\frac{\omega}{\gamma}\right) \right]_0^{+\infty}} \\
&= \frac{1}{\frac{\pi}{2}(N-1)} \\
&= \frac{2}{(N-1)\pi},
\end{aligned}$$

so that

$$\begin{aligned}
\rho_{\mathcal{E}_N}(\omega, \gamma) &= \frac{K\gamma(N-1)}{\gamma^2 + \omega^2} \\
&= \frac{2\gamma}{\pi(\gamma^2 + \omega^2)}.
\end{aligned} \tag{9}$$

For the fully connected network we have

$$\begin{aligned}
\rho_{\mathcal{F}_N}(\omega, \gamma) &= K \sum_{i=1}^{N-1} \frac{\gamma}{\gamma^2 + (\omega - \omega_i)^2} \\
&= K \sum_{i=1}^{N-1} \frac{\gamma}{\gamma^2 + (\omega - \sqrt{N})^2} \\
&= \frac{\gamma K(N-1)}{\gamma^2 + \omega^2 + N - 2\omega\sqrt{N}},
\end{aligned}$$

where  $K$  is

$$\begin{aligned}
K &= \frac{1}{\gamma(N-1) \int_0^{+\infty} \frac{d\omega}{\gamma^2 + \omega^2 + N - 2\omega\sqrt{N}}} \\
&= \frac{1}{\frac{\gamma(N-1)}{\gamma} \left[ \arctan\left(\frac{\omega - \sqrt{N}}{\gamma}\right) \right]_0^{+\infty}} \\
&= \frac{1}{(N-1) \left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right)},
\end{aligned}$$

so that

$$\begin{aligned}
\rho_{\mathcal{F}_N}(\omega, \gamma) &= \frac{\gamma K(N-1)}{\gamma^2 + \omega^2 + N - 2\omega\sqrt{N}} \\
&= \frac{1}{(N-1) \left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right)} \cdot \frac{\gamma(N-1)}{\gamma^2 + \omega^2 + N - 2\omega\sqrt{N}} \\
&= \frac{\gamma}{\left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} .
\end{aligned} \tag{10}$$

Thus, we expand Eq. 6 as follows:

$$\begin{aligned}
1 &= \epsilon_\gamma(\mathcal{E}_N, \mathcal{F}_N) \\
&= \sqrt{\int_0^\infty (\rho_{\mathcal{E}_N}(\omega, \gamma) - \rho_{\mathcal{F}_N}(\omega, \gamma))^2 d\omega} \\
&= \sqrt{\int_0^\infty \left( \frac{2\gamma}{\pi(\gamma^2 + \omega^2)} - \frac{\gamma}{\left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} \right)^2 d\omega} \\
&= \sqrt{\int_0^\infty A^2 d\omega + \int_0^\infty B^2 d\omega - 2 \int_0^\infty AB d\omega} ,
\end{aligned} \tag{11}$$

where

$$\begin{aligned}
A &= \frac{2\gamma}{\pi(\gamma^2 + \omega^2)} \\
B &= \frac{\gamma}{\left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} .
\end{aligned}$$

The three terms in Eq. 11 can be expanded as follows:

$$\begin{aligned}
\int_0^{+\infty} A^2 d\omega &= \int_0^{+\infty} \left( \frac{2\gamma}{\pi(\gamma^2 + \omega^2)} \right)^2 d\omega \\
&= \frac{4\gamma^2}{\pi^2} \int_0^{+\infty} \frac{d\omega}{(\gamma^2 + \omega^2)^2} \\
&= \frac{4\gamma^2}{\pi^2} \frac{1}{2\gamma^3} \left[ \frac{\gamma\omega}{\gamma^2 + \omega^2} + \arctan\left(\frac{\omega}{\gamma}\right) \right]_0^{+\infty} \\
&= \frac{2}{\gamma\pi^2} \left[ \frac{\pi}{2} \right] \\
&= \frac{1}{\pi\gamma} ;
\end{aligned} \tag{12}$$



$$\begin{aligned}
\int_0^{+\infty} B^2 d\omega &= \int_0^{+\infty} \left( \frac{\gamma}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} \right)^2 d\omega \\
&= \int_0^{+\infty} \frac{\gamma^2}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right)^2 (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})^2} d\omega \\
&= \frac{\gamma^2}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right)^2} \int_0^{+\infty} \frac{d\omega}{(\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})^2} \\
&= \frac{\gamma^2}{2\gamma^3 \left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right)^2} \left[ \frac{\gamma(\omega - \sqrt{N})}{\gamma^2 + (\omega - \sqrt{N})^2} + \arctan\left(\frac{\omega - \sqrt{N}}{\gamma}\right) \right]_0^{+\infty} \\
&= \frac{1}{2\gamma \left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right)^2} \left( \frac{\pi}{2} + \frac{\gamma\sqrt{N}}{\gamma^2 + N} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) ; \tag{13}
\end{aligned}$$

$$\begin{aligned}
-2 \int_0^{+\infty} AB d\omega &= -2 \int_0^{+\infty} \frac{2\gamma}{\pi(\gamma^2 + \omega^2)} \frac{\gamma}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} d\omega \\
&= \frac{-2 \cdot \gamma \cdot 2\gamma}{\pi \left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right)} \int_0^{+\infty} \frac{d\omega}{(\gamma^2 + \omega^2) (\gamma^2 + \omega^2 + N - 2\omega\sqrt{N})} \\
&= \frac{-4\gamma}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right) \pi(4\gamma^2 + N)} \left[ \frac{\gamma}{\sqrt{N}} \log \frac{\gamma^2 + \omega^2}{\gamma^2 + \omega^2 + N - 2\omega\sqrt{N}} + \right. \\
&\quad \left. \arctan\left(\frac{\omega - \sqrt{N}}{\gamma}\right) + \arctan\left(\frac{\omega}{\gamma}\right) \right]_0^{+\infty} \\
&= \frac{-4\gamma}{\left(\frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right)\right) \pi(4\gamma^2 + N)} \left[ \frac{\pi}{2} + \frac{\pi}{2} - \frac{\gamma}{\sqrt{N}} \log \frac{\gamma^2}{\gamma^2 + N} + \right. \\
&\quad \left. \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right] . \tag{14}
\end{aligned}$$

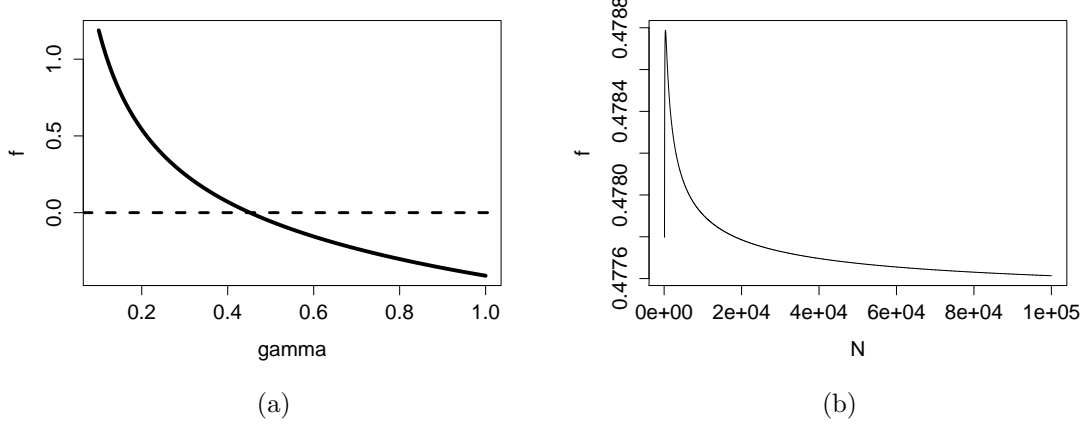


Figure 12: (a) Behaviour of  $f(\gamma, 10)$  and solution of Eq. 6; (b) Behaviour of  $f(\bar{\gamma}, N)$

Plugging Eqs. 12,13,14 into Eq. 11, we obtain:

$$\begin{aligned}
1 &= \epsilon_\gamma(\mathcal{E}_N, \mathcal{F}_N) \\
&= \frac{1}{\pi\gamma} + \frac{1}{2\gamma \left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right)^2} \left( \frac{\pi}{2} + \frac{\gamma\sqrt{N}}{\gamma^2 + N} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) - \\
&\quad \frac{-4\gamma}{\left( \frac{\pi}{2} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right) \pi(4\gamma^2 + N)} \left[ \pi - \frac{\gamma}{\sqrt{N}} \log \frac{\gamma^2}{\gamma^2 + N} + \arctan\left(\frac{\sqrt{N}}{\gamma}\right) \right]. \tag{15}
\end{aligned}$$

Consider now the function  $f(N, \gamma) = \epsilon_\gamma(\mathcal{E}_N, \mathcal{F}_N) - 1$ : for a fixed value of  $N$ , it is a monotonically decreasing function of  $\gamma$ , so the equation Eq. 6 has an unique solution  $\bar{\gamma}$ . In Fig. 12(a) we display the situation for  $N = 10$ . As a function of  $N$ , the behaviour of  $f(N, \bar{\gamma})$  is not monotonic, but its range is rather narrow, as exemplified in Fig. 12(b).

## B Uniqueness of $\bar{\gamma}^\dagger$

The spectra of the laplacian matrices of the two extremal graphs  $\hat{\mathcal{E}}_N^\dagger$  and  $\hat{\mathcal{F}}_N^\dagger$  are now

$$\text{spec}(L(\hat{\mathcal{E}}_N^\dagger)) = \underbrace{(0, \dots, 0)}_{2N} \quad \text{and} \quad \text{spec}(L(\hat{\mathcal{F}}_N^\dagger)) = (0, \underbrace{N-2, \dots, N-2}_{N-1}, \underbrace{N, \dots, N}_{N-1}, 2N-2).$$

It follows that

$$K_{\hat{\mathcal{E}}_N^\dagger} = \frac{2}{(2N-1)\pi}$$

and

$$K_{\hat{\mathcal{F}}_N^\dagger} = \frac{1}{(2N-1)\frac{\pi}{2} + (N-1) \left( \arctan \frac{\sqrt{N-2}}{\gamma} + \arctan \frac{\sqrt{N}}{\gamma} \right) + \arctan \frac{\sqrt{2N-2}}{\gamma}}.$$

Thus the equation

$$\epsilon_\gamma(\hat{\mathcal{E}}^\dagger, \hat{\mathcal{F}}^\dagger) = 1$$

(whose solution is the normalizing factor  $\bar{\gamma}^\dagger$ ) reads as follows:

$$1 = \sqrt{\int_0^{+\infty} \left[ \frac{2\gamma}{\gamma^2 + \omega^2} - \frac{\gamma \left( \frac{N-1}{\gamma^2 + (\omega - \sqrt{N-2})^2} + \frac{N-1}{\gamma^2 + (\omega - \sqrt{N})^2} + \frac{1}{\gamma^2 + (\omega - \sqrt{2N-2})^2} \right)}{(2N-1)\frac{\pi}{2} + (N-1) \left( \arctan \frac{\sqrt{N-2}}{\gamma} + \arctan \frac{\sqrt{N}}{\gamma} \right) + \arctan \frac{\sqrt{2N-2}}{\gamma}} \right]^2 d\omega}. \quad (16)$$

Introduce now a few shorthands: define, for  $T, U \in \mathbb{R}$ , the following integral

$$\int_0^{+\infty} \frac{d\omega}{(\gamma^2 + (\omega - \sqrt{T})^2)(\gamma^2 + (\omega - \sqrt{U})^2)} = \begin{cases} M(T) & \text{if } T = U, \\ L(T, U) & \text{if } T \neq U. \end{cases}$$

Then,

$$M(T) = \frac{\frac{1}{2} \left( \gamma^2 \arctan \frac{\sqrt{T}}{\gamma} + T \arctan \frac{\sqrt{T}}{\gamma} + \gamma \sqrt{T} \right)}{\gamma^5 + T\gamma^3} + \frac{\pi}{4\gamma^3},$$

and

$$L(T, U) = \frac{-\log(\gamma^2 + U) + \log(\gamma^2 + T)}{(4\gamma^2 + T + 3U)\sqrt{T} - (4\gamma^2 + 3T + U)\sqrt{U}} + \frac{\pi + \arctan\left(\frac{\sqrt{T}}{\gamma}\right) + \arctan\left(\frac{\sqrt{U}}{\gamma}\right)}{4\gamma^3 + T\gamma - 2\sqrt{T}\sqrt{U}\gamma + U\gamma}.$$

To shorten notations, define furthermore

$$Z = \frac{2\gamma}{\pi} \quad W = \gamma(N-1)K_{\hat{\mathcal{F}}_N^\dagger} \quad W' = \frac{W}{N-1}.$$

With the aforementioned positions, Eq. 16 becomes

$$\begin{aligned} 1 = & Z^2 M(0) + W^2 M(N-2) + W^2 M(N) + W'^2 M(2N-2) \\ & - 2ZWL(0, N-2) - 2ZWL(0, N) - 2ZW'L(0, 2N-2) \\ & + 2W^2 L(N-2, N) + 2WW'L(N-2, 2N-2) + 2WW'L(N, 2N-2). \end{aligned} \quad (17)$$

As in the undirected case, for each  $N$  Eq. 17 has an unique solution  $\bar{\gamma}^\dagger$ , whose value is quite close to  $\bar{\gamma}$ , as shown in Tab. B.

N	$\bar{\gamma}$	$\bar{\gamma}^\dagger$
5	0.4272836	0.3866861
10	0.4517012	0.4300291
50	0.4752742	0.4704579
100	0.4777976	0.4753463
500	0.4787492	0.4782538
1000	0.4785596	0.4783119
10000	0.4779060	0.4778813

Table 1: Comparison of  $\bar{\gamma}$  and  $\bar{\gamma}^\dagger$  for different number of nodes  $N$ .